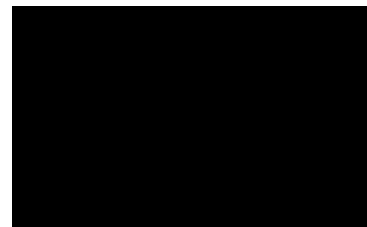



# Statistical Analysis Plan



---


<b>Sponsor Name:</b>	Jazz Pharmaceuticals
<b>Protocol Number:</b>	14-003
<b>Protocol Title:</b>	A Twelve-week, Double-blind, Placebo-controlled, Randomized, Parallel-group, Multicenter Study of the Safety and Efficacy of JZP-110 [(R)-2-amino-3-phenylpropylcarbamate hydrochloride] in the Treatment of Excessive Sleepiness in Subjects with Obstructive Sleep Apnea (OSA)
<b>Protocol Version and Date:</b>	Original protocol, 17 December 2014 Protocol Amendment 1, 18 February 2015 Protocol Amendment 2, 10 September 2015 Protocol Amendment 3, 08 February 2016

<b>SAP Version:</b>	1.0
<b>SAP Version Date:</b>	25 Jan 2017

---

## Notice of Confidential and Proprietary Information:

The information contained in this document is confidential belonging to Jazz Pharmaceuticals. Acceptance of this document constitutes agreement by the recipient that no information contained herein will be published or disclosed without prior written authorization from an official of Jazz Pharmaceuticals. However, this document may be disclosed to appropriate Institutional Review Board and Ethics Committees or duly authorized representatives of a national regulatory authority under the condition that they are requested to keep it confidential. In the event of an actual or suspected breach of this obligation,  should be notified promptly.

## Statistical Analysis Plan

**Version:** 1.0

**Version Date:** 25 Jan 2016

I confirm that I have reviewed this document and agree with the content.

APPROVALS	
Written by: [REDACTED]	Date
Approved by: [REDACTED]	Date
<i>Jazz Pharmaceuticals</i>	
Approved by: [REDACTED]	Date
Approved by: [REDACTED]	Date

## Statistical Analysis Plan

### TABLE OF CONTENTS

<b>1. ABBREVIATION .....</b>	<b>7</b>
<b>2. PURPOSE .....</b>	<b>10</b>
<b>3. STUDY OBJECTIVES .....</b>	<b>11</b>
3.1. Primary Objective .....	11
3.2. Secondary Objective(s) .....	11
3.3. Study Design .....	11
3.4. Determination of Sample Size .....	16
3.5. Randomization and Stratification.....	16
3.6. Administration of Study Drug.....	16
<b>4. ENDPOINTS .....</b>	<b>17</b>
4.1. Co-Primary Efficacy Endpoints .....	17
4.2. Key Secondary Efficacy Endpoint .....	17
4.3. Other Secondary Efficacy Endpoints .....	17
4.4. Functional Outcomes and Quality of Life Endpoints .....	17
4.5. Exploratory Endpoints .....	18
4.6. Pharmacokinetic Endpoints .....	18
4.7. Pharmacodynamic Endpoints .....	18
4.8. Safety Endpoints.....	19
<b>5. STATISTICAL HYPOTHESES FOR TRIAL OBJECTIVES .....</b>	<b>20</b>
<b>6. ANALYSIS POPULATIONS .....</b>	<b>21</b>
6.1. Safety Population.....	21

## Statistical Analysis Plan

6.2.	Modified Intent-to-Treat Population .....	21
6.3.	Per Protocol Population .....	21
6.4.	Pharmacokinetic Population .....	21
7.	<b>EFFICACY ASSESSMENTS</b> .....	<b>22</b>
7.1.	Maintenance of Wakefulness Test (MWT) .....	22
7.2.	Epworth Sleepiness Scale (ESS) .....	22
7.3.	Clinician Global Impression of Severity (CGIs) .....	22
7.4.	Clinician Global Impression of Change (CGIc) .....	23
7.5.	Patient Global Impression of Change (PGIc) .....	23
7.6.	Functional Outcomes of Sleep Questionnaire Short Version (FOSQ-10) .....	23
7.7.	36-Item Short Form Health Survey Version 2 (SF-36v2) .....	23
7.8.	EuroQoL EQ-5D-5L .....	23
7.9.	Work Productivity and Activity Impairment Questionnaire: Specific Health Problem V2.0 (WPAI:SHP) .....	24
7.10.	Primary OSA Therapy Use .....	24
8.	<b>GENERAL ASPECTS FOR STATISTICAL ANALYSIS</b> .....	<b>25</b>
8.1.	General Methods .....	25
8.2.	<b>KEY Definitions</b> .....	<b>25</b>
8.2.1.	Baseline .....	25
8.2.2.	Study Day .....	25
8.2.3.	JZP-110 Exposure .....	26
8.3.	Visit Windows .....	26
8.4.	Missing Data .....	27
8.5.	Level of Significance and Multiplicity Adjustment .....	28
8.6.	Pooling of Centers and Region .....	29

## Statistical Analysis Plan

<b>8.7.</b>	<b>Subgroups and Subgroups Analyses.....</b>	<b>29</b>
<b>9.</b>	<b>DEMOGRAPHIC, OTHER BASELINE CHARACTERISTICS AND MEDICATION .....</b>	<b>31</b>
<b>9.1.</b>	<b>Subject Disposition and Withdrawals.....</b>	<b>31</b>
<b>9.2.</b>	<b>Demographic and Other Baseline Characteristics .....</b>	<b>31</b>
<b>9.3.</b>	<b>Medical/Surgical History .....</b>	<b>31</b>
<b>9.4.</b>	<b>Medication/Procedure .....</b>	<b>32</b>
9.4.1.	Prior Medication.....	32
9.4.2.	Concomitant Medication and Procedure .....	32
<b>9.5.</b>	<b>Protocol Deviations .....</b>	<b>32</b>
<b>10.</b>	<b>EFFICACY .....</b>	<b>34</b>
<b>10.1.</b>	<b>Co-Primary Efficacy Endpoints and Analyses.....</b>	<b>34</b>
10.1.1.	Primary Analysis .....	34
10.1.2.	Secondary Analysis .....	35
10.1.3.	Sensitivity Analyses .....	35
10.1.4.	Subgroup Analyses .....	40
<b>10.2.</b>	<b>Key Secondary Efficacy Endpoint and Analyses .....</b>	<b>40</b>
10.2.1.	Primary Analysis .....	40
10.2.2.	Secondary Analysis .....	40
10.2.3.	Sensitivity Analyses .....	40
10.2.4.	Subgroup Analyses .....	41
<b>10.3.</b>	<b>Other Secondary Efficacy Endpoints and Analyses .....</b>	<b>41</b>
10.3.1.	Time Course of Efficacy on the MWT: Change in Sleep Latency Time (in Minutes) on Each of the 5 MWT Trials .....	41
10.3.2.	CGIc: Percentage of Subjects Reported as Improved (Minimally, Much, or Very Much) at Weeks 1, 4, 8, and 12.....	41
10.3.3.	MWT: Change in the Mean Sleep Latency Time (in Minutes) as Determined from the First Four Trials of a 40-minute MWT from Baseline to Week 4 .....	42
10.3.4.	ESS: Change in ESS Score from Baseline to Weeks 1, 4, and 8 .....	42
10.3.5.	PGIc: Percentage of Subjects Reported as Improved (Minimally, Much, or Very Much) at Weeks 1, 4, and 8.....	42
<b>10.4.</b>	<b>Functional Outcomes and Quality of Life Endpoints .....</b>	<b>42</b>
10.4.1.	FOSQ-10: Change in the Total Score and 5 Subscales from Baseline to Weeks 1, 4, 8, and 12 .....	42
10.4.2.	SF-36v2: Change in 8 Domain Scores, Physical Component Summary (PCS) Score, and Mental Component Summary (MCS) Score from Baseline to Weeks 4, 8, and 12 .....	43

## Statistical Analysis Plan

10.4.3. EQ-5D-5L: EQ-5D Dimensions at Week 1, 4, 8, and 12 .....	44
10.4.4. EQ VAS: Change in the VAS Score from Baseline to Weeks 1, 4, 8, and 12.....	44
10.4.5. EQ-5D-5L Index: Index Value from Baseline to Week 1, 4, 8, and 12 .....	44
10.4.6. WPAI:SHP: Change in WPAI:SHP Endpoints from Baseline to Week 1,4,8, and 12 .....	45
<b>10.5. Exploratory Endpoints .....</b>	<b>46</b>
10.5.1. Use of Primary OSA Therapy .....	46
10.5.2. Change in PSG Parameters from Baseline to Week 4 and 12 .....	47
10.5.3. Categorical Analysis for Maintenance of Wakefulness Test .....	48
<b>10.6. Summary of Efficacy Analysis Methods.....</b>	<b>48</b>
<b>11. ANALYSIS OF PHARMACOKINETICS .....</b>	<b>51</b>
<b>12. SAFETY .....</b>	<b>52</b>
12.1. Extent of Exposure .....	52
12.2. Treatment Compliance .....	52
12.3. Adverse Events .....	52
12.3.1. Subgroup Analyses .....	53
12.4. Laboratory Evaluations .....	53
12.5. Vital Signs and 24-Hour Ambulatory Blood Pressure Monitoring .....	54
12.5.1. Regular Vital Signs .....	54
12.5.2. Blood Pressure and Pulse on the MWT day .....	55
12.5.3. 24 hour Ambulatory Blood Pressure Monitoring (ABPM) .....	56
12.6. ECG .....	56
12.7. Columbia-Suicide Severity Rating Scale.....	57
12.7.1. C-SSRS Outcomes/Composite Scores .....	57
12.7.2. Analyses .....	58
<b>13. INTERIM ANALYSES .....</b>	<b>59</b>
<b>14. CHANGE FROM ANALYSIS PLANNED IN PROTOCOL .....</b>	<b>60</b>
<b>15. REFERENCE .....</b>	<b>61</b>

## Statistical Analysis Plan

### 1. ABBREVIATION

Abbreviation	Description
AE	Adverse Event
AHI	Apnea Hypopnea Index
ANCOVA	Analysis of Covariance
AI	Apnea Index
ATC	Anatomical Therapeutic Chemical
BMI	Body Mass Index
CGIc	Clinical Global Impression of Change
CGIs	Clinical Global Impression of Severity
CI	Confidence Interval
CRF	Case Report Form
CSR	Clinical Study Report
C-SSRS	Columbia-Suicide Severity Rating Scale
CV	Coefficient of Variation
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
ECG	Electrocardiogram
ESS	Epworth Sleepiness Scale
ET	Early Termination
EQ VAS	EuroQoL Visual Analog Scale
FOSQ-10	Functional Outcomes of Sleep Questionnaire Short Version
ICSD-3	International Classification of Sleep Disorders, 3rd Edition
IVRS	Interactive Voice Response System
IWRS	Interactive Web Response System
LOCF	Last Observation Carried Forward
LOE	Lack of Efficacy
LS	Least Square
MAR	Missing at Random

## Statistical Analysis Plan

Abbreviation	Description
Max	Maximum
MCAR	Missing Completely at Random
MCS	Mental Component Summary
MedDRA	Medical Dictionary for Regulatory Activities
MI	Multiple Imputation
Min	Minimum
mITT	Modified Intent-to-Treat
MMRM	Mixed-Effect Model with Repeated Measures
MNAR	Missing Not at Random
MWT	Maintenance of Wakefulness Test
OSA	Obstructive Sleep Apnea
PGIc	Patient Global Impression of Change
PAP	Positive airway pressure
PCS	Physical Component Summary
PK	Pharmacokinetics
PP	Per-Protocol
PSG	Polysomnography
QTcF	QT Corrected with Fridericia's Formula
SAE	Serious Adverse Event
SAP	Statistical Analysis Plan
SD	Standard Deviation
SF-36v2	36-Item Short Form Health Survey Version 2
SI	Single Imputation
TEAE	Treatment Emergent Adverse Event
TST	Total Sleep Time
WASO	Wake after Sleep Onset
WHO	World Health Organization
WPAI:SHP	Work Productivity and Activity Impairment Questionnaire:



Jazz Pharmaceuticals  
Protocol: 14-003  
Version 1.0

## Statistical Analysis Plan

Abbreviation	Description
	Specific Health Problem

## Statistical Analysis Plan

---

### 2. PURPOSE

This Statistical Analysis Plan (SAP) is created based on Protocol 14-003 Amendment 3 (08 February 2016) and it describes in detail the statistical methodology and the statistical analyses to be conducted for the above mentioned protocol.

Results obtained from the analyses outlined in this document will become the basis of the final clinical study report (CSR) for this protocol. The purpose of this plan is to provide specific instructions as to how each analysis will be conducted. Any deviations from these guidelines must be substantiated by sound statistical reasoning and documented in the final CSR.

## Statistical Analysis Plan

### 3. STUDY OBJECTIVES

#### 3.1. PRIMARY OBJECTIVE

To evaluate the efficacy of JZP-110 administered once daily for up to 12 weeks in daily doses of 37.5, 75, 150, and 300 mg compared to placebo in the treatment of excessive sleepiness in adult subjects with OSA.

#### 3.2. SECONDARY OBJECTIVE(S)

To evaluate the safety and tolerability of JZP-110 administered once daily for up to 12 weeks in daily doses of 37.5, 75, 150, and 300 mg compared to placebo in the treatment of excessive sleepiness in adult subjects with OSA.

To characterize the pharmacokinetics (PK) of JZP-110 in subjects with OSA using sparse sampling methods

#### 3.3. STUDY DESIGN

The Schedule of Events is presented in Table 1, and the Study Schema in Figure 1.

This trial is a 12-week, randomized, double-blind, placebo-controlled, multicenter, 5-arm parallel group study of safety and efficacy of JZP-110 in the treatment of excessive sleepiness in adult subjects with OSA. Following the successful completion of Screening and Baseline visits, stratified randomization on the basis of subjects' compliant or non-compliant use of their primary OSA therapy will be used to assign subjects in a 1:1:2:2:2 ratio to JZP-110 37.5, 75, 150, or 300 mg daily, or placebo. Treatment will be administered once daily over a Treatment Phase of 12 weeks. Subjects randomized to the 150 mg daily dose will initially receive 75 mg daily from Day 1 through Day 3 of the first week of the Treatment Phase and will receive 150 mg daily starting on Day 4. Subjects randomized to the 300 mg daily dose will initially receive 150 mg daily from Day 1 through Day 3 of the first week of the Treatment Phase and will receive 300 mg daily starting on Day 4. Subjects randomized to the other treatment groups will not require titration.

During the Treatment Phase, subjects will return to the investigative site to complete efficacy and safety assessments at the end of Weeks 1, 4, 8, and 12; the Week 4 and 12 visits will also include an overnight stay at the investigational site for nocturnal polysomnography (PSG) followed by a Maintenance of Wakefulness Test (MWT), and the Week 8 Visit will include 24-hour ambulatory blood pressure monitoring. Subjects will take their final dose of study drug at the Week 12 visit prior to the Week 12 visit assessments. Subjects will return at the end of Week 14 for follow-up assessments.

## Statistical Analysis Plan

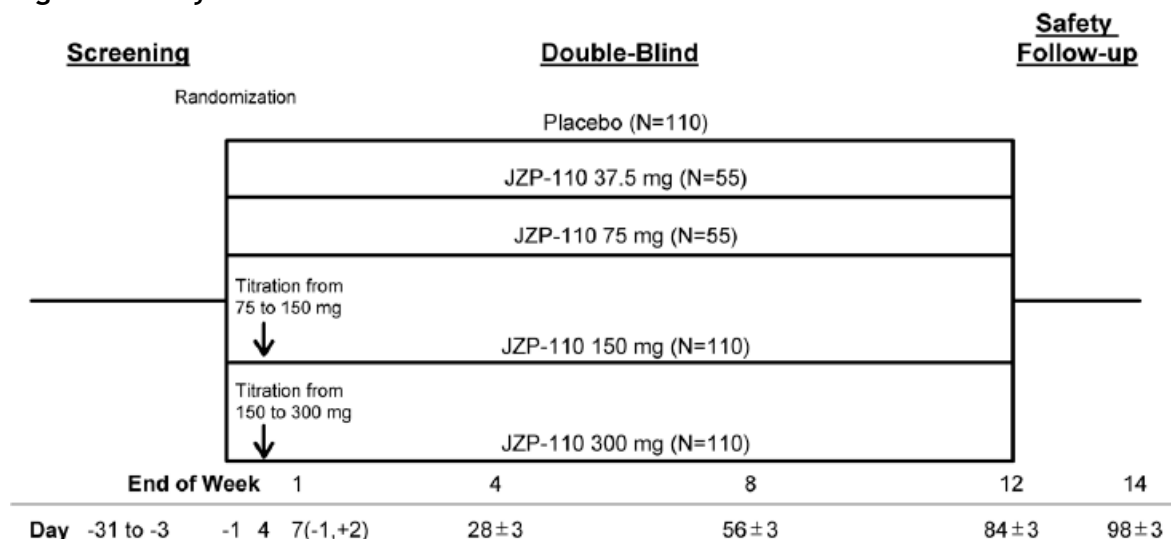
Unless there are any outstanding safety issues that require follow-up, subjects will be discharged from the study at the Week 14 visit.

Efficacy will be assessed by changes in the mean sleep latency on the first 4 trials of a 5-trial, 40-minute MWT and by changes in the mean Epworth Sleepiness Scale (ESS) score as co-primary endpoints, percentage of patients improved on the Patient Global Impression of Change (PGIc) as a key secondary endpoint, and Clinical Global Impression of Change (CGIc), Functional Outcomes of Sleep Questionnaire Short Version (FOSQ-10), 36-Item Short Form Health Survey version 2 (SF-36v2), EuroQoL EQ-5D-5L, and the Work Productivity and Activity Impairment Questionnaire: Specific Health Problem V2.0 (WPAI:SHP) as other endpoints.

Four blood samples will be collected from each subject for PK evaluations: one sample during the Week 1 visit, one sample during the Week 4 visit, and two samples during the Week 8 visit. The samples collected at the Week 1 and Week 8 visits will be within 1-8 hours of dosing. The sample collected at the Week 4 visit will be within 8-12 hours of dosing.

Safety will be assessed by the incidence of observed and reported adverse events (AEs), and changes in physical examination findings, electrocardiograms (ECGs), clinical laboratory tests, vital signs, 24-hour ambulatory blood pressure monitoring, and the Columbia-Suicide Severity Rating Scale (C-SSRS). Safety will be assessed throughout the study.

**Figure 1 Study Schema**



## Statistical Analysis Plan

Table 1 Schedule of events

	Screening	Baseline	Treatment Phase												Early Term	Safety Follow-up <sup>a</sup>
Visit	1	2	3	4	5	6	7	8	9	10	11	12	13	14		15
Day/ End of Week	Day -31 to -3	Day -2&-1	Day 7 (-1, +2) Wk 1	Day 14 (±3) Wk 2	Day 21 (±3) Wk 3	Day 27&28 (±3) Wk 4	Day 35 (±3) Wk 5	Day 42 (±3) Wk 6	Day 49 (±3) Wk 7	Day 56 (±3) Wk 8	Day 63 (±3) Wk 9	Day 70 (±3) Wk 10	Day 77 (±3) Wk 11	Day 83&84 (±3) Wk 12		Day 98 (±3) Wk 14
Clinic visit	X	X	X			X				X				X	X	X
Phone Contact				X	X		X	X	X		X	X	X			
Informed consent	X															
Inclusion/ Exclusion	X	X														
Demographics	X															
Medical history	X	X														
Physical examination	X													X	X	
Height	X															
Weight	X	X	X			X				X				X	X	X
Vital signs	X	X	X			X				X				X	X	X
ECG	X	X	X			X				X				X	X	X
Fasting serum chemistry, hematology, urinalysis	X	X <sup>b</sup>												X	X	
Urine sample for possible drug screen	X	X <sup>c</sup>	X			X <sup>c</sup>				X <sup>c</sup>				X	X	
Serum pregnancy test	X															
Urine pregnancy test		X												X	X	
Light breakfast	X	X				X								X	X	
Light lunch		X				X								X	X <sup>d</sup>	

## Statistical Analysis Plan

	Screening	Baseline	Treatment Phase												Early Term	Safety Follow-up <sup>a</sup>
Visit	1	2	3	4	5	6	7	8	9	10	11	12	13	14		15
Day/ End of Week	Day -31 to -3	Day -2&-1	Day 7 (-1, +2) Wk 1	Day 14 (±3) Wk 2	Day 21 (±3) Wk 3	Day 27&28 (±3) Wk 4	Day 35 (±3) Wk 5	Day 42 (±3) Wk 6	Day 49 (±3) Wk 7	Day 56 (±3) Wk 8	Day 63 (±3) Wk 9	Day 70 (±3) Wk 10	Day 77 (±3) Wk 11	Day 83&84 (±3) Wk 12		Day 98 (±3) Wk 14
Clinic visit	X	X	X			X				X				X	X	X
Phone Contact				X	X		X	X	X		X	X	X			
C-SSRS	X	X	X			X				X				X	X	X
Concomitant medication	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Review and remind to record primary OSA therapy use	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Instruct how to discontinue excluded medications	X															
Dispense 24-hr ambulatory blood pressure monitor	X									X						
Overnight stay PSG/MWT		X <sup>c</sup>				X								X	X <sup>d</sup>	
ESS		X	X			X				X				X	X <sup>d</sup>	
FOSQ-10		X	X			X				X				X	X <sup>d</sup>	
SF-36v2		X				X				X				X	X <sup>d</sup>	
EQ-5D-5L		X	X			X				X				X	X <sup>d</sup>	
WPAI-SHP		X	X			X				X				X	X <sup>d</sup>	
CGIs		X														
Randomization		X														
Dispense study drug		X	X			X				X						
Collect study drug/ assess compliance			X			X				X				X	X	
PGIc			X			X				X				X	X <sup>d</sup>	
CGIc			X			X				X				X	X <sup>d</sup>	

## Statistical Analysis Plan

	Screening	Baseline	Treatment Phase												Early Term	Safety Follow-up <sup>a</sup>
Visit	1	2	3	4	5	6	7	8	9	10	11	12	13	14		15
Day/ End of Week	Day -31 to -3	Day -2&-1	Day 7 (-1, +2) Wk 1	Day 14 (±3) Wk 2	Day 21 (±3) Wk 3	Day 27&28 (±3) Wk 4	Day 35 (±3) Wk 5	Day 42 (±3) Wk 6	Day 49 (±3) Wk 7	Day 56 (±3) Wk 8	Day 63 (±3) Wk 9	Day 70 (±3) Wk 10	Day 77 (±3) Wk 11	Day 83&84 (±3) Wk 12		Day 98 (±3) Wk 14
Clinic visit	X	X	X			X				X				X	X	X
Phone Contact				X	X		X	X	X		X	X	X			
Study drug administration prior to MWT			X			X								X	X <sup>d</sup>	
Blood draws for PK			X			X				X						
Assess interest in Study 14-005										X				X		
Adverse Events		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Schedule next clinic visit and/or phone contact	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	



Shaded columns indicate clinic visits.

- The 2 week Safety Follow-up is not required for subjects who enter the Open-label Safety Study (14-005) at the Final Clinic Visit.
- If the Baseline visit occurs outside of the screening window (i.e., >29 days after the Screening Visit) obtain blood samples for serum chemistry, hematology and a urine sample for urinalysis. Record if the subject was fasting or nonfasting at the time of the collection.
- Samples collected at Screening, Baseline, and Week 12 or Early Termination will be analyzed. Samples collected at other visits may be analyzed at the investigator's discretion.
- These assessments are not required for subjects who are withdrawing early from the study and are unable or unwilling to take additional study drug to complete a final PSG/MWT.
- For subjects who are being rescreened, a repeat PSG and MWT is not required at Visit 2 if the results of the previous PSG and MWT meet the current inclusion/exclusion criteria and there have been no changes to the medical history, concomitant medications, or primary OSA therapy that would likely affect the MWT.

## Statistical Analysis Plan

### 3.4. DETERMINATION OF SAMPLE SIZE

Approximately 440 subjects are planned for enrollment with approximately 110 subjects per treatment group in the placebo, 150 mg, and 300 mg groups. A sample size of 99 subjects per group (placebo, 150 mg, and 300 mg groups) will provide at least 90% power to detect a difference of 5 minutes in the mean sleep latency time as determined from the MWT (mean of the first four trials) and a difference of 3.5 points on the ESS changes from Baseline to Week 12 between each of these JZP-110 treatment dose groups and placebo. This calculation assumes common standard deviations of 10 minutes for the MWT and 6 points for the ESS changes from Baseline and a two-sided significance level of 0.05 using a t-test. To account for dropouts without evaluable data, a sample size of 110 subjects per treatment group (placebo, 150 mg, and 300 mg groups) is planned. Approximately 55 subjects will be randomized to each of the JZP-110 37.5 mg and 75 mg groups.

### 3.5. RANDOMIZATION AND STRATIFICATION

All subjects will be randomly assigned to treatments via an interactive voice response system (IVRS) or an Interactive Web Response System (IWRS). Subjects will be randomized to JZP-110 37.5 mg, 75 mg, 150 mg, 300 mg or placebo in a 1:1:2:2:2 ratio using stratified randomization. The strata are compliant or non-compliant use of primary OSA therapy (compliant versus non-compliant) at screening.

### 3.6. ADMINISTRATION OF STUDY DRUG

Subjects will receive JZP-110 37.5, 75, 150, or 300 mg or placebo once daily as a single oral dose over a Treatment Phase of 12 weeks.



## Statistical Analysis Plan

### 4. ENDPOINTS

#### 4.1. CO-PRIMARY EFFICACY ENDPOINTS

- MWT: Change in the mean sleep latency time (in minutes) as determined from the first four trials of a 40-minute MWT from Baseline to Week 12
- ESS: Change in ESS score from Baseline to Week 12

#### 4.2. KEY SECONDARY EFFICACY ENDPOINT

- PGlc: Percentage of subjects reported as improved (minimally, much, or very much) at Week 12

#### 4.3. OTHER SECONDARY EFFICACY ENDPOINTS

- Time course of efficacy on the MWT: Change in sleep latency time (in minutes) on each of the 5 MWT trials at Week 4, and Week 12
- CGlc: Percentage of subjects reported as improved (minimally, much, or very much) at Week 12
- MWT: Change in the mean sleep latency time (in minutes) as determined from the first four trials of a 40-minute MWT from Baseline to Week 4
- ESS: Change in ESS score from Baseline to Week 1, Week 4, and Week 8
- PGlc: Percentage of subjects reported as improved at Week 1, Week 4, and Week 8
- CGlc: Percentage of subjects reported as improved at Week 1, Week 4, and Week 8

#### 4.4. FUNCTIONAL OUTCOMES AND QUALITY OF LIFE ENDPOINTS

- FOSQ-10: Change in the total score and 5 FOSQ-10 subscales from Baseline to Week 1, Week 4, Week 8, and Week 12
- SF-36v2: Change in the physical component summary (PCS) and the mental component summary (MCS) and change in the 8 subscales from Baseline to Week 4, Week 8, and Week 12
- EQ-5D-5L:

## Statistical Analysis Plan

- EQ-5D Dimensions:
  - Number and percentage of subjects in each of the 5 levels (e.g., no problem, slight problem, moderate problem, severe problem, unable) for each dimension (e.g., mobility, self-care) over time
  - Number and percentage of subjects reporting any problems (levels 2-5) for each dimension (e.g., mobility, self-care) over time
- EuroQoL Visual Analog Scale (EQ VAS): Mean and SD or median with 25th and 75th percentiles for the VAS at Baseline, Week 1, Week 4, Week 8 and Week 12. Change in the VAS scores from Baseline to Week 1, Week 4, Week 8, and Week 12
- EQ-5D-5L Index: Index value at Baseline to Week 1, Week 4, Week 8, and Week 12
- WPAI:SHP: Percent work time missed due to problem over time, percent impairment while working due to problem over time, percent overall work impairment due to problem over time, and percent activity impairment due to problem over time

### 4.5. EXPLORATORY ENDPOINTS

- Change in the percentage of use of primary OSA therapy from Baseline to each subsequent analysis period (e.g., Weeks 1-4, Weeks 5-8, and Weeks 9-12)
- Change in PSG parameters including total sleep time (TST), time in Stages N1, N2, N3, wake after sleep onset (WASO), number of awakenings, apnea index (AI), Apnea Hypopnea Index (AHI), number of central apneas, SaO<sub>2</sub> nadir, and SaO<sub>2</sub> mean from Baseline to Week 4, and Week 12

### 4.6. PHARMACOKINETIC ENDPOINTS

- Concentration of JZP-110 at Weeks 1, 4 and 8.

### 4.7. PHARMACODYNAMIC ENDPOINTS

Not applicable

## Statistical Analysis Plan

---

### 4.8. SAFETY ENDPOINTS

To evaluate safety and tolerability as determined by the occurrence of and/or changes in:

- Treatment-emergent adverse events (TEAEs)
- Change in clinical laboratory tests (chemistry, hematology, and urinalysis)
- Vital signs
- 24 hour ambulatory blood pressure monitoring
- 12-lead electrocardiograms (ECGs)
- C-SSRS

## Statistical Analysis Plan

### 5. STATISTICAL HYPOTHESES FOR TRIAL OBJECTIVES

The two co-primary hypotheses, corresponding to the two co-primary efficacy variables (MWT and ESS), are the following:

- JZP-110 is superior to placebo as measured by change in the mean sleep latency time (in minutes) as determined from the first four trials of a 40-minute MWT from Baseline to Week 12

The statistical null hypothesis is that for the MWT, the mean change in the mean sleep latency time (in minutes) from Baseline to Week 12 for each JZP-110 dose group is the same as the mean change in the mean sleep latency time (in minutes) from Baseline to Week 12 for placebo group.

- JZP-110 is superior to placebo as measured by change in ESS score from Baseline to Week 12

The statistical null hypothesis is that for the ESS score, the mean change from Baseline to Week 12 for each JZP-110 dose group is the same as the mean change from Baseline to Week 12 for the placebo group.

In addition to the two co-primary hypotheses, the key secondary hypothesis is:

- JZP-110 is superior to placebo as measured by Patient Global Impression of Change (PGIc) at Week 12

The statistical null hypothesis is that for the PGIc, the proportion of subjects reported as improved (minimally, much, or very much) at Week 12 for each JZP-110 dose group is the same as the proportion of subjects reported as improved (minimally, much, or very much) at Week 12 for the placebo group.

## Statistical Analysis Plan

### **6. ANALYSIS POPULATIONS**

#### **6.1. SAFETY POPULATION**

The Safety Population will consist of all subjects who received at least one dose of study medication. This population will be used for safety analyses and will be presented in the tables and listings of safety data.

#### **6.2. MODIFIED INTENT-TO-TREAT POPULATION**

The Modified Intent-to-Treat (mITT) Population will include subjects who were randomized, received at least one dose of study medication and have baseline and at least one post-baseline evaluation of MWT or ESS. If a subject in the mITT Population does not have an assessment for a particular endpoint, that subject will be excluded in the analysis of that endpoint.

The mITT population will be used for efficacy and functional outcomes and quality of life analyses.

#### **6.3. PER PROTOCOL POPULATION**

The Per-Protocol (PP) population will include subjects who completed the trial according to protocol specifications without a major protocol violation. Based on the protocol deviation management plan, a major violation can be identified by an exclusion from Per Protocol population flag in the study protocol deviation report. The reasons for exclusion will also be detailed in the study protocol deviation report.

The categories of major protocol deviations will be defined in the protocol deviation management plan, and the plan will be approved before unblinding of the study.

The PP population will be used in a secondary analysis of the co-primary endpoints and the key secondary endpoint.

#### **6.4. PHARMACOKINETIC POPULATION**

The PK Population will include subjects who received at least one dose of JZP-110 and have evaluable PK data for the PK analysis.

## Statistical Analysis Plan

### **7. EFFICACY ASSESSMENTS**

#### **7.1. MAINTENANCE OF WAKEFULNESS TEST (MWT)**

The MWT is the standard objective measure of an individual's ability to remain awake during the daytime in a darkened, quiet environment and is commonly used to assess response to treatment. A five-trial, 40 minute MWT will be performed at Baseline and at the Week 4 and 12 visits (or Early Termination if the subject is willing and able to take study drug for the assessments) on the morning after an overnight PSG according to a standard protocol, which will be provided in a manual to each site.

Each MWT during the study should be started at approximately the same time of the day. During the MWT trials, subjects should be seated in bed in a darkened room with their back and head supported by a bedrest such that their neck is not uncomfortably flexed or extended. Subjects will be instructed to sit still and remain awake for as long as possible during each of the 5 40-minute trials separated by 2-hour intervals. Following a light breakfast, the subject will be allowed to relax prior to initiating the first MWT trial. The first MWT trial should occur approximately 2 hours after "lights on" at the Baseline visit and approximately 1 hour after dosing with study drug at all other visits at which the MWT will be conducted. If the subject falls asleep during a trial, they will be awakened and instructed to remain awake until the next trial. If the subject does not fall asleep, then the specific trial is terminated at 40 minutes and a sleep latency of 40 minutes is recorded. The subject is then instructed to remain awake (and will be awoken if they fall asleep) until the next trial.

#### **7.2. EPWORTH SLEEPINESS SCALE (ESS)**

The ESS is a self-administered questionnaire with 8 questions asking the subject how likely they would be to doze off or fall asleep in different situations. Responses range from 0 = would never doze to 3 = high chance of dozing (Protocol Appendix 5). Subjects will be asked to complete the ESS with regard to the level of sleepiness they experienced over the past 7 days at the Baseline and Week 1, 4, 8, and 12 visits (or Early Termination). It provides a measure of a person's general level of daytime sleepiness, or their average sleep propensity in daily life. The ESS is a validated measure with high specificity and sensitivity for assessing subjective sleepiness.

#### **7.3. CLINICIAN GLOBAL IMPRESSION OF SEVERITY (CGIs)**

The CGIs is a 7-point Likert-type rating scale and a widely used assessment in clinical psychopharmacology trials to assess severity of illness (Protocol Appendix 12). The responses of this investigator-completed scale range from 1 = normal, no signs of illness to 7 = among the most extremely ill patients. The Investigator will rate his/her impression of the severity of the subject's current condition at Baseline relative to his/her experience with this patient population at Baseline.

## Statistical Analysis Plan

### **7.4. CLINICIAN GLOBAL IMPRESSION OF CHANGE (CGIc)**

The CGIc is a 7-point Likert-type rating scale and a widely used assessment to assess efficacy in clinical drug trials. Investigators will rate their impression of any change in the subject's condition from baseline (before the subject started treatment) on a 7-point scale ranging from 1 = very much improved to 7 = very much worse at the Week 1, 4, 8, and 12 visits (or Early Termination) (Protocol Appendix 13).

### **7.5. PATIENT GLOBAL IMPRESSION OF CHANGE (PGIc)**

The PGIc is a 7-point Likert-type rating scale and a widely used measure to assess efficacy in clinical drug trials. Subjects will rate the change in their condition since they started treatment on a 7-point scale ranging from 1 = very much improved to 7 = very much worse at the Week 1, 4, 8, and 12 visits (or Early Termination) (Protocol Appendix 14).

### **7.6. FUNCTIONAL OUTCOMES OF SLEEP QUESTIONNAIRE SHORT VERSION (FOSQ-10)**

The FOSQ is a 30-item disease specific quality of life questionnaire to determine functional status in adults; measures are designed to assess the impact of disorders of excessive sleepiness on multiple activities of everyday living and the extent to which these activities are improved by effective treatment. The FOSQ-10 is a short version of the original 30-item FOSQ that has been shown to perform similarly to the longer version. The FOSQ-10 has been shown to exhibit high internal consistency, effect sizes, and pre- and post-treatment differences that are highly correlated with the longer version. Subjects will complete the FOSQ-10 at the Baseline and Week 1, 4, 8, and 12 visits (or Early Termination) (Protocol Appendix 6).

### **7.7. 36-ITEM SHORT FORM HEALTH SURVEY VERSION 2 (SF-36v2)**

The SF-36v2 is a multi-purpose, short-form health survey with 36 questions. It yields an 8-scale profile of functional health and well-being scores as well as psychometrically-based physical and mental health summary measures and a preference-based health utility index. Subjects will complete the SF-36v2 at the Baseline and Week 4, 8, and 12 visits (or Early Termination) (Protocol Appendix 7).

### **7.8. EuroQoL EQ-5D-5L**

The EQ-5D-5L is a standardized instrument for use as a measure of health outcome that includes a descriptive system consisting of 5 dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) and an EQ visual analogue scale (VAS). It is applicable to a wide range of health conditions and treatments. It provides a simple descriptive profile and a single index value for health status. The EQ-5D-5L

## Statistical Analysis Plan

includes five levels of severity for each of the 5 dimensions of the descriptive system and was developed to improve the instrument's reliability and sensitivity and to reduce ceiling effects. Subjects will complete the EQ-5D-5L at the Baseline and Week 1, 4, 8, and 12 visits (or Early Termination) (Protocol Appendix 8).

### **7.9. WORK PRODUCTIVITY AND ACTIVITY IMPAIRMENT QUESTIONNAIRE: SPECIFIC HEALTH PROBLEM V2.0 (WPAI:SHP)**

The WPAI:SHP questionnaire is a 6-item self-administered questionnaire that measures work time missed and work and activity impairment because of a specified health problem during the past 7 days. The WPAI:SHP will be used with "OSA" as the specified health problem. The validity of the WPAI has been established in a number of diseases (Reilly et al. 1993). Subjects will complete the WPAI:SHP at the Baseline and Week 1, 4, 8, and 12 visits (or Early Termination) (Protocol Appendix 10).

### **7.10. PRIMARY OSA THERAPY USE**

For the purposes of this study, primary OSA therapy is defined as use of PAP, oral pressure therapy, an oral appliance, or an upper airway stimulator. Subjects who report using a primary OSA therapy will have information regarding whether they used their device each night and the duration of nightly use extracted from the data download from their device or memory card at each clinic visit from Screening and through the Final Visit at Week 12. Subjects who report using a device for which usage data cannot be retrieved will record their primary OSA therapy usage and the estimated duration of use (more than half of the night, less than half of the night, or don't know) on a daily basis from Screening and through the Final Visit at Week 12. Subjects who report not using a primary OSA therapy at Screening will be asked to confirm that they have continued to not use a primary OSA therapy. The study staff will review the information that each subject provides regarding their primary OSA therapy use at each study visit and will discuss it with the subject at each phone contact. A stable level of use of a primary OSA therapy is defined as a change of <30% of the number of nights used per week.

Subjects will be encouraged to stay on their current primary OSA therapy at the same level of use throughout the study.



## Statistical Analysis Plan

### 8. GENERAL ASPECTS FOR STATISTICAL ANALYSIS

#### 8.1. GENERAL METHODS

Unless otherwise specified, for numeric data, descriptive statistics will include the number of subjects with data to be summarized (n), mean, standard deviation (SD), median, minimum (min) and maximum (max). The same number of decimal places as in the raw data will be presented when reporting min and max, 1 more decimal place than in the raw data will be presented when reporting mean and median, and 2 more decimal places than in the raw data will be presented when reporting SD. If the raw data have 3 decimals or more, 3 decimals will be presented for mean, median, min and max, and SD.

All categorical/qualitative data will be presented using absolute and relative frequency counts and percentage. All percentages will be presented with one-decimal point. Percentages equal to 100 will be presented as 100% and percentages will not be presented for zero frequencies, but the categories whose counts are zero will be displayed for the sake of completeness.

P-value > 0.9999 will be presented as '>0.9999' and p-value < 0.0001 will be presented as '< 0.0001'.

All analyses and summary outputs will be generated by treatment group (Placebo, JZP-110 37.5 mg, JZP-110 75 mg, JZP-110 150 mg, JZP-110 300 mg and combined JZP-110) using SAS® version 9.3 (or higher).

All data collected in this study will be presented in by-subject listings.

#### 8.2. KEY DEFINITIONS

##### 8.2.1. Baseline

The Baseline measurement for a variable is defined as the last non-missing value from the baseline visit measured prior to the first dose of the study drug. If a subject has repeated measurements from the baseline visit, then the last repeated non-missing value will be used. If there is not a value from the scheduled baseline visit, the last non-missing value from other screening or unscheduled visits measured prior to the first dose of study drug will be used.

##### 8.2.2. Study Day

A study day will be assigned as follows:

- The first dose of study drug is designated as Day 1.

## Statistical Analysis Plan

- For visit days after Day 1, study day = visit date - Day 1 date + 1.
- For visit days prior to Day 1, study day = visit date - Day 1 date. Thus, study days for screening visit are negative numbers. There is no “Day 0”.
- The end date of treatment is the last dose date in the study.
- The end date of the study is defined as the date of the subject’s last assessment including the safety follow-up in the study.

### 8.2.3. JZP-110 Exposure

JZP-110 exposure is defined as the number of days that a dose of JZP-110 is taken, regardless of the JZP-110 dose level.

### 8.3. VISIT WINDOWS

The unscheduled or early termination (ET) visit will be mapped to a scheduled visit for analysis using the date of collection/assessment as a basis to determine study day and then study day will be mapped to the intended visit. The table below contains the visit windows.

Once analysis visit windows are assigned, all visits, including scheduled visits, unscheduled visits, and ET visits will be eligible for being flagged as the “analyzed record” within the analysis visit window, a subject’s individual analysis visit window could potentially contain more than one visit. In the event of multiple visits falling within an analysis visit window or in case of a tie, the following rules will be used in sequence to determine the “analyzed record” for the analysis visit window:

- If there is a scheduled visit/week for the analysis visit window, then the scheduled visit/week data will be used.
- If there is no scheduled visit/week for the analysis visit window, the data closest to the scheduled day will be used.
- If there is no scheduled visit/week for the analysis visit window and there is a tie between the data in the number of days before and after the scheduled day, the later data will be used.

## Statistical Analysis Plan

The unscheduled or early termination (ET) visit will be mapped to a scheduled visit before the imputation methods will be used for handling dropouts and missing data for efficacy endpoints.

The data not flagged as the “analyzed record” will also be listed in subject listings.

Study Day Window	Scheduled day	Scheduled Visit/Week
Days (-31) to (-3)	Day -3	Visit 1/Screening
Day (-2) to (-1)	Day -2 & -1	Visit 2/Baseline
Day 1 - 10	Day 7	Visit 3/Week 1
Day 11 - 17	Day 14	Visit 4/Week 2
Day 18 - 24	Day 21	Visit 5/Week 3
Day 25 - 31	Day 27 & 28	Visit 6/Week 4
Day 32 - 38	Day 35	Visit 7/Week 5
Day 39 - 45	Day 42	Visit 8/Week 6
Day 46 - 52	Day 49	Visit 9/Week 7
Day 53 - 59	Day 56	Visit 10/Week 8
Day 60 - 66	Day 63	Visit 11/Week 9
Day 67 - 73	Day 70	Visit 12/Week 10
Day 74 - 80	Day 77	Visit 13/Week 11
Day 81 - 87	Day 83 & 84	Visit 14/Week 12
Day > 87	Day 98	Visit 15/Week 14

### 8.4. MISSING DATA

For the analysis of the co-primary efficacy parameters of MWT and the total ESS score, missing data will be handled by mixed-effect model with repeated measure (MMRM) for the primary analysis. For sensitivity analyses, the following missing data approaches will be used to impute the missing data to assess the potential impact of missing data: two single imputation (SI) approaches [last observation carried forward (LOCF) and mean imputation] as described in section 10.1.3.1 and two multiple imputation (MI) approaches [MCMC with regression method and Pattern Mixture model using dropout pattern imputation method] as described in 10.1.3.2.

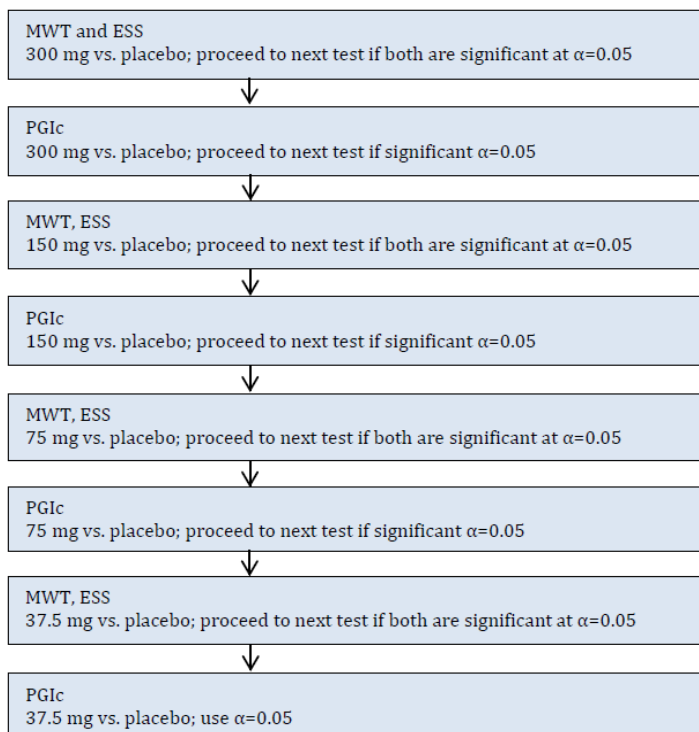
For the key secondary efficacy parameter of PGlc and the other secondary efficacy parameter of CGlc, missing data will be imputed by the last observation carried forward (LOCF) for the primary analyses. Two SI approaches (worst case and varies by early termination reason) will be used to impute the missing data to assess potential impact of missing data as sensitivity analyses. See Section 10.2.1 and 10.2.3.

## Statistical Analysis Plan

### 8.5 LEVEL OF SIGNIFICANCE AND MULTIPLICITY ADJUSTMENT

To address the multiplicity issue of multiple endpoints and dose groups, the fixed hierarchical testing sequence in Figure 2 will be used. Testing will begin with the comparison of 300 mg daily dose versus placebo for the co-primary efficacy endpoints MWT and ESS. Since they are co-primary endpoints, both have to be significant at the 0.05 level from the primary analysis before the test can proceed to the next level (PGIc at 300 mg daily dose versus placebo). Testing will proceed from the highest JZP-110 dose to the lowest dose and will stop when a significance level exceeds 0.05 (see Figure 2). This gate keeping approach will control the family-wise error rate at 0.05 for the comparisons of the four JZP-110 doses versus placebo in MWT, ESS, and PGIc. Although the study is not powered for the 37.5 and 75 mg treatment arms, the testing will proceed to levels as outlined in Figure 2. The dose(s) that show(s) significant difference versus placebo in both MWT and ESS will be considered efficacious doses and additional testing to characterize the time course of efficacy will be performed (Section 10.3.1).

**Figure 2. Multiplicity Strategy**



## Statistical Analysis Plan

---

### 8.6. POOLING OF CENTERS AND REGION

Because of the large number of centers and the fact that many centers have a small number of subjects, analyses will not be performed by center and will not include an adjustment for center. Data from all investigational centers will be pooled.

Data will also be summarized by region (North America and Europe) for specified analyses.

### 8.7. SUBGROUPS AND SUBGROUPS ANALYSES

Exploratory analyses of the key efficacy and safety endpoints will be conducted in the following subgroups of subjects:

- Subgroups of subjects who report compliant and non-compliant use of a primary OSA therapy
- Subgroups of subjects by region (North America and Europe)
- Subgroups of subjects by country (e.g., US, Canada, France, Germany, Netherlands)

The following endpoints will be included in the subgroup analyses:

- MWT: Change in the mean sleep latency time (in minutes) as determined from the first four trials of a 40-minute MWT from Baseline to each post-baseline time point
- ESS: Change in ESS score from Baseline to each post-baseline time point
- PGIC: Percentage of subjects reported as improved at each time point
- CGIC: Percentage of subjects reported as improved at each time point
- TEAEs
- EQ-5D-5L:
  - Number and percentage of subjects in each of the 5 levels (e.g., no problem, slight problem, moderate problem, severe problem, unable) for each dimension at Baseline and each post-baseline time point
  - Number and percentage of subjects reporting any problems (levels 2-5) for each dimension at Baseline and each post-baseline time point
  - EQ VAS score at each time point and change from Baseline to each post-baseline time point

EQ-5D-5L index value at each time point and change from Baseline to each post-baseline time point

- FOSQ-10: Change in total score and 5 subscales from Baseline to each post-baseline time point

## Statistical Analysis Plan

---

For subgroups of subjects by country, only descriptive statistics will be performed.

## Statistical Analysis Plan

### **9. DEMOGRAPHIC, OTHER BASELINE CHARACTERISTICS AND MEDICATION**

#### **9.1. SUBJECT DISPOSITION AND WITHDRAWALS**

Number of all enrolled subjects and number and percentage of subjects in each analysis population will be presented. In addition, number of subjects who completed/prematurely discontinued, and reason for discontinuation will also be presented by treatment group for each analysis population and subgroup of subjects by region.

The summary of disposition over time will show the number of subjects terminating the study in each week.

For screen failure subjects, reasons for screen failure will be summarized separately.

#### **9.2. DEMOGRAPHIC AND OTHER BASELINE CHARACTERISTICS**

Demographics and baseline characteristics will be summarized by treatment group for the Safety Population, mITT Population, Per-Protocol Population, PK Population, and subgroups of subjects by region.

Demographics and baseline characteristics include: age, gender, race, ethnicity, region, country, height, weight and body mass index (BMI).

- Height (in cm) = height (in inches) \* 2.54
- Weight (in kg) = weight (in lbs) \* 0.4536

For screen failure subjects, demographics will be summarized separately.

In addition, the following variables will be summarized:

- Baseline disease severity: mean MWT sleep latency time, total ESS score, and clinical global impression of severity (CGIs)
- Randomization stratification factor (compliant or non-compliant use of their primary OSA) as collected on the CRF

#### **9.3. MEDICAL/SURGICAL HISTORY**

Medical/surgical history collected at screening and baseline will be coded using the Medical Dictionary for Regulatory Activities (MedDRA), Version 18.

## Statistical Analysis Plan

Medical/surgical history data will be summarized by system organ class (SOC) and preferred term (PT) and treatment group for the safety population.

### 9.4. MEDICATION/PROCEDURE

Medications will be coded using WHO DRUG dictionary, Version: March, 2015.

Procedures will be coded using MedDRA, Version 18.0.

#### 9.4.1. Prior Medication

Prior medications will be defined as medications with a start date prior to the first dose of study drug. The stop date of the medication should be before or after the first dose of study drug or the medication may be ongoing. If a start date is completely missing, then the medication will be considered a prior medication.

Prior medications will be summarized by anatomical therapeutic chemical (ATC) level 3 term and preferred term based on the safety population.

#### 9.4.2. Concomitant Medication and Procedure

Concomitant medications will be defined as medications with a stop date on or after the first dose of study drug or any medication that is ongoing. The start date of the medication may be before or after the first dose of study drug. A medication with completely missing use dates or partially missing use dates without evidence that the medication was stopped prior to the first dose of study drug will be considered a concomitant medication. Concomitant procedures will be defined as procedures with a procedure date on or after first dose of study drug. A procedure with a completely or partially missing use date without evidence that the procedure was conducted prior to the first dose of study drug will be considered a concomitant procedure.

Concomitant medications will be summarized by ATC level 3 term, preferred term based on the safety population. Procedures will be summarized by MedDRA SOC and PT. The summaries will be based on the safety population.

### 9.5. PROTOCOL DEVIATIONS

Before database lock, the cumulative protocol deviation report will be generated by the clinical team and reviewed by the Jazz Medical Monitor and Lead Statistician per protocol deviation management plan. Source data will include both deviations reported directly from the sites and deviations identified from CRF data through programmatic edit checks. A flag of exclusion from the Per Protocol population due to major



## Statistical Analysis Plan

violation(s) will be identified and the reasons for exclusion from the Per Protocol population will be documented based on discussion between the Jazz Medical Monitor, Lead Statistician, and Clinical Operations Lead. The flag of exclusion from the Per Protocol population and the reason for exclusion from the Per Protocol population will be included in the cumulative protocol deviation report and the final protocol deviation report. After database lock, the protocol deviation report will be filed in the TMF and will be used as source data for the clinical study database.

Major protocol deviations will be summarized based on the safety and mITT populations.

## Statistical Analysis Plan

### 10. EFFICACY

Observed data at each time point and change from baseline to each post-baseline visit/time point will be summarized by treatment group. For categorical data, frequency counts and percentages will be presented in a similar manner.

#### 10.1. CO-PRIMARY EFFICACY ENDPOINTS AND ANALYSES

The co-primary efficacy endpoints are:

- MWT: Change in the mean sleep latency time (in minutes) as determined from the first four trials of a 40-minute MWT from Baseline to Week 12
- ESS: Change in ESS score from Baseline to Week 12

The mean sleep latency time at each time point is the average of the sleep latency time (in minutes) from the first four trials of a 40-minute MWT at the specific time point. If two or more of the first four individual trials MWT sleep latencies are missing at a given time point, the mean sleep latency on the MWT will be set to missing at the specific time point. If one of the first four individual trials MWT sleep latency is missing, the mean of the remaining three of the first four MWT trials will be used for calculating the mean sleep latency on the MWT at that specific time point.

The MWT sleep latency can range from 0 to 40 minutes, with higher latencies indicating greater ability to stay awake. Thus, an increase in mean sleep latency on the MWT from Baseline represents an improvement in the ability to stay awake.

The ESS total score is the sum of eight item scores. If three or more item scores are missing at a specific time point, the ESS total score will be set to missing. If one or two ESS items are missing at specific time point, the mean of the remaining seven or six non-missing ESS items at that time point will be used to impute the missing ESS items. The ESS total score will be calculated as the sum of the observed and the imputed item scores.

The ESS score can range from 0 to 24 points, with higher scores indicating a greater chance of dozing. Thus, a decrease in ESS score represents an improvement in excessive sleepiness.

##### 10.1.1. Primary Analysis

For both co-primary efficacy endpoints, a mixed-effect model with repeated measures (MMRM) will be used as the primary method of analysis. The response variable will be the change from Baseline to each post-baseline visit and the co-primary endpoints will be the changes from Baseline to Week 12. The model will include the following fixed effects:

## Statistical Analysis Plan

- Treatment (with five levels: JZP-110 37.5 mg, JZP-110 75 mg, JZP-110 150 mg, JZP-110 300 mg, and placebo for the primary analysis or with 2 levels: Combined JZP-110 and placebo for the additional exploratory analysis )
- Visit (as a discrete and repeated factor)
- Treatment-by-visit interaction
- Baseline of the respective efficacy endpoint, a continuous covariate
- Randomization stratification factor (with 2 levels, compliant or non-compliant use of a primary OSA therapy based on CRF data)

“Baseline” refers to the baseline value of the response variable (mean MWT sleep latency time or total ESS score). SAS procedure PROC MIXED will be used to carry out this analysis. All available data will be included in the model. An unstructured covariance matrix will be used to model the correlation among repeated measurements unless convergence fails; in this case an auto-regressive (AR1) structure will be used. If the AR1 structure fails to converge, a compound symmetric structure will be used. The relative Hessian convergence criterion will be used in the MMRM model. Least square (LS) estimates of treatment differences (each JZP-110 dose group versus placebo for the primary analysis and combined JZP-110 group versus placebo for the additional exploratory analysis) and their 95% confidence intervals will be presented respectively.

The primary analysis will be based on the mITT population.

### 10.1.2. Secondary Analysis

For both co-primary efficacy endpoints, the analysis will be performed using the same statistical model as the primary analysis based on the PP population, instead of the mITT population.

### 10.1.3. Sensitivity Analyses

This section describes analyses to explore the potential impact of missing data. Importantly, the analyses described in Section 10.1.1 are anticipated to be the most appropriate methods; the additional methods described in this section will be used to qualitatively evaluate the robustness of the primary analysis method.

The sensitivity analyses will be based on the mITT population.

#### 10.1.3.1. Sensitivity Analyses using Single Imputation (SI) Approach

The analysis of covariance (ANCOVA) model will be used to evaluate the change in co-primary endpoints at Week 12 using the SI approaches described below. The ANCOVA model will include treatment as a fixed effect and the baseline value of the endpoint and randomization stratification factor as covariates.

## Statistical Analysis Plan

- **SI Approach 1:** Last observation carried forward

SI approach 1 is a simple and common method for handling missing data. LOCF is used extensively in spite of the fact that it requires the following two strong assumptions: (1) missing data due to dropouts completely at random (MCAR) following early discontinuation, and (2) response following a subject dropping out remains constant following the last observed value prior to drop out. For each subject, missing data at Week 12 will be replaced with the last non-missing value.

- **SI Approach 2:** Mean Imputation

SI approach 2 also assumes MCAR, meaning that the probability of an observed data point being missing does not depend either on the observed data or unobserved data. For each subject, missing data at Week 12 will be replaced with the corresponding treatment group mean.

### 10.1.3.2. Sensitivity Analyses using Multiple Imputation (MI) Approach

Instead of filling in a single value for each missing value, Rubin's (1987) multiple imputation procedure replaces each missing value with a set of plausible values that represent the uncertainty about the right value to impute.

Collected data are often incomplete, with a mixed monotone and non-monotone structure. For example, some subjects may drop out from the study, resulting in a monotone pattern of missing data. Alternatively, data may be missing intermittently due to invalid measurements; this non-monotone pattern is considered missing at random (MAR).

Two MI methods are planned:

**Markov chain Monte Carlo (MCMC) method (Schafer 1997):** For non-monotone missing data patterns, the MCMC assumes multivariate normality and generates multiple imputations using simulations from a Bayesian prediction distribution for normal data. Without prior information about the mean and covariance estimates, a noninformative prior will be used under the MCMC method.

**Regression method:** For monotone missing data patterns, a regression model is fitted for each variable with missing values. Based on the resulting model, a new regression model is simulated to impute subsequent missing values for the variable (Rubin 1987, pp.166-167). The process is repeated sequentially for variables with missing values.

Based on these two MI methods, the missing data will be imputed using the following approaches.

## Statistical Analysis Plan

### **MI Approach 1: Multiple Imputation using MCMC and Regression Method**

Step 1: non-monotone missing data will be imputed using the MCMC method within each treatment group

Step 2: monotone missing data in each treatment group will be imputed using the regression method on the basis of the predicted future pattern for the same treatment group

### **MI Approach 2: Pattern Mixture Model using Dropout Pattern Imputation**

Pattern-Mixture Model using dropout pattern imputation is planned to explore the possibility of non-ignorable missing data.

For the missing data due to dropout, pattern-mixture imputation will be considered based on the dropout patterns.

Scenario 1:

MNAR (missing not at random):

- dropout due to lack of efficacy (LOE) in JZP-110 treatment groups

MAR (missing at random):

- dropout due to other reason (not LOE ) in JZP-110 treatment groups
- dropout due to any reason in the placebo group

Scenario 2:

MNAR (missing not at random):

- dropout due to AE in JZP-110 treatment groups

MAR (missing at random):

- dropout due to other reason (not AE ) in JZP-110 treatment groups
- dropout due to any reason in the placebo group

Scenario 3:

MNAR (missing not at random):

## Statistical Analysis Plan

- dropout due to AE or LOE in JZP-110 treatment groups

MAR (missing at random):

- dropout due to other reason (not AE or LOE ) in the treatment groups
- dropout due to any reason in the placebo group

A tipping point method will be used to investigate the robustness of the primary analysis method. Imputed values in JZP-110 treatment groups that fall into a MNAR pattern will be adjusted under different scenarios above (Carpenter and Kenward 2013, pp. 237-239; van Buuren 2012, pp. 88-89) by using the delta adjustment imputation method.

If missing data in the JZP-110 treatment groups falls into a MNAR pattern, the analysis will assume the treatment differences at the primary analysis visit (Week 12) progressively decrease from 0%, 10%, 20%, .., and up to 100% (i.e., equivalent to placebo). Otherwise, MCMC for the non-monotone missing data based on MAR assumption and regression method for monotone missing data on the basis of the predicted future pattern for the same treatment group will be applied.

The tipping point analysis will be performed for each co-primary endpoint separately if the primary analysis results are statistically significant ( $p\text{-value} < 0.05$ ). The procedure will be implemented using the steps delineated below:

Step 1: non-monotone missing data in JZP-110 treatment and placebo groups will be imputed using the MCMC method within each treatment group.

Step 2: the monotone missing data in the placebo group and JZP-110 treatment groups that falls into MAR pattern will be imputed using the regression method on the basis of the predicted future pattern for the same treatment group.

Step 3: for the monotone missing data in the JZP-110 treatment groups (e.g., 37.5mg, 75 mg, 150 mg, and 300 mg) that fall into MNAR pattern:

- 3.1 For missing data at the visits (e.g., Week 1, Week 4, or Week 8) before the primary analysis visit at Week 12 (e.g, Week 1, Week 4, or Week 8):

Based on the observed data in the MNAR subgroup (e.g., AE, LOE, and (AE or LOE)), the following regression methods will be used in sequence to determine the applicable “imputation method” to impute the MNAR data prior to the primary analysis visit at Week 12:

## Statistical Analysis Plan

- The regression method on the basis of the predicted future pattern for the same subgroup (e.g., AE, LOE, or (AE or LOE)) by the JZP-110 treatment groups (37.5mg, 75 mg, 150 mg, 300mg)
  - The regression method on the basis of the predicated future pattern for the same subgroup (e.g., AE, LOE, or (AE or LOE)) while combining the JZP-110 treatment groups
  - The regression method on the basis of the predicated future pattern for the early termination subgroup regardless of the ET reasons while combining the JZP-110 treatment groups
  - The regression method on the basis of the predicted future pattern for the same treatment group, which is the same approach at step 2.
- 3.2 For the primary analysis visit (i.e. Week 12), missing data at Week 12 will be imputed by regression method by using the imputed data prior to the Week 12 (Section 3.1) and all of the observed data from the treatment group
- 3.3 The imputed value at the primary analysis visit (i.e. Week 12) for JZP-110 treatment groups that fall into the MNAR pattern will be subtracted progressively by a delta =  $k$  (0%, 10%, 20%, .., 100%) \* LS mean treatment difference at this visit (i.e. Week 12) obtained from the primary analysis method described in Section 10.1.1.

Depending on the number of observed data in the JZP-110 treatment groups that fall into an MNAR pattern, step 3 will be performed for the three drop out scenarios separately: dropout due to LOE in the JZP-110 treatment groups, dropout due to AE in the JZP-110 treatment groups, and drop out due to LOE or AE in JZP-110 treatment groups.

Step 4: One hundred (100) imputed datasets will be generated for each MI analysis. Each imputed dataset is analyzed separately using the MMRM model specified in Section 10.1.1. The final estimate of treatment difference will be the average of the estimates based on the 100 individual imputed datasets. The pooling of the individual estimates and inferences based on the combined estimate will be handled by SAS procedure MIANALYZE.

Step 3.3 will be repeated iteratively while increasing the penalty (e.g., 10%, 20%, ..., 100%) for the missing data in the JZP-110 treatment groups that fall into an MNAR pattern and keeping all the other imputed data unchanged, until the tipping point value (i.e., where  $p\text{-value} > 0.05$ ) is identified.

## Statistical Analysis Plan

### 10.1.4. Subgroup Analyses

The primary analysis in Section 10.1.1 will be used for the subgroup analyses. However, the randomization stratification factor (with 2 levels, compliant or non-compliant use of a primary OSA therapy based on the CRF data) will be removed from the model for analysis by subgroups of subjects with compliant or non-compliant use of primary OSA therapy at randomization. The analysis will be based on the mITT population.

### 10.2. KEY SECONDARY EFFICACY ENDPOINT AND ANALYSES

The key secondary efficacy endpoint is percentage of subjects reported as improved (minimally, much, or very much) on the PGlc at Week 12.

#### 10.2.1. Primary Analysis

The percentage of subjects reported as improved on the PGlc at Week 12 will be calculated and summarized by treatment group. Comparisons will be performed between each JZP-110 group vs placebo for primary analysis and combined JZP-110 vs placebo for additional explorative analysis using a chi-square test. 95% confidence intervals for the difference in proportions will be calculated. Missing data at Week 12 for the key secondary efficacy endpoint will be imputed using LOCF. The analysis will be based on the mITT population.

#### 10.2.2. Secondary Analysis

The secondary analysis of the key secondary endpoint will be performed using the same statistical method as the primary analysis based on the PP population, instead of mITT population.

#### 10.2.3. Sensitivity Analyses

Missing data at Week 12 will be imputed using the following single imputation approaches:

##### Approach 1 - SI Varies by Early Termination Reason

Subjects with missing data at Week 12 due to discontinuation by lack of efficacy or adverse events will be considered non-responders at Week 12. Subjects with missing data at Week 12 for other reasons will be imputed using LOCF.

##### Approach 2 - SI by the Worst-Case

Subjects with missing data at Week 12 will be considered non-responders.



## Statistical Analysis Plan

The chi-square test will be used to test treatment difference between each JZP-110 dose group vs placebo and combined JZP-110 vs placebo. 95% confidence intervals for the difference in the percentages between treatment groups will be calculated. These analyses will be based on the mITT population.

### 10.2.4. Subgroup Analyses

The primary analysis in Section 10.2.1 will be used for the subgroup analyses. Analyses will be based on the mITT population.

## 10.3. OTHER SECONDARY EFFICACY ENDPOINTS AND ANALYSES

Other secondary efficacy endpoints will be analyzed based on the mITT population.

### 10.3.1. Time Course of Efficacy on the MWT: Change in Sleep Latency Time (in Minutes) on Each of the 5 MWT Trials

For the time course of efficacy analysis, the dose(s) that show(s) significant difference versus placebo in both MWT and ESS will be considered efficacious dose(s). For each of these doses, additional testing to characterize the time course of efficacy on the MWT will be performed. The mean and standard error of the mean MWT data will be displayed graphically for the dose and placebo. Pairwise comparison vs. placebo for each of the 5 MWT trials will be conducted at a significance level of 0.05 using a similar MMRM model described in Section 10.1.1. Characterizing the time course of efficacy will begin with identifying the first trial that demonstrates significant difference from placebo. If such a trial is identified, the treatment difference at the next trial will be examined. If the treatment difference for the next trial is also significant at the 0.05 level, the treatment will be considered efficacious at the time of the next trial. This procedure will continue as long as significance for each subsequent trial is observed or until trial 5 is examined.

### 10.3.2. CGIc: Percentage of Subjects Reported as Improved (Minimally, Much, or Very Much) at Weeks 1, 4, 8, and 12

The percentage of subjects reported as improved on the CGIc at Week 1, 4, 8, and 12 will be calculated and summarized by treatment group. Comparisons will be performed between each JZP-110 group vs placebo and combined JZP-110 vs placebo using a chi-square test at each time point; 95% confidence intervals for the difference in proportions will be calculated. LOCF will be used for subjects with missing data.

Similar analyses will be conducted for the subgroups listed in Section 8.7.

## Statistical Analysis Plan

### **10.3.3. MWT: Change in the Mean Sleep Latency Time (in Minutes) as Determined from the First Four Trials of a 40-minute MWT from Baseline to Week 4**

Change in mean sleep latency time (in minutes) as determined from the first four trials of a 40-minute MWT from Baseline to Week 4 will be analyzed using the methods described for the primary analysis of the co-primary efficacy endpoints (Section 10.1.1). Similar analyses will be conducted for the subgroups listed in Section 8.7.

### **10.3.4. ESS: Change in ESS Score from Baseline to Weeks 1, 4, and 8**

Change in ESS score from Baseline to Weeks 1, 4, and 8 will be analyzed using the methods described for the primary analysis of the co-primary efficacy endpoints (Section 10.1.1).

Similar analyses will be conducted for the subgroups listed in Section 8.7.

### **10.3.5. PGIc: Percentage of Subjects Reported as Improved (Minimally, Much, or Very Much) at Weeks 1, 4, and 8**

The percentage of subjects reported as improved at Weeks 1, 4, and 8 will be analyzed using the methods described for the primary analysis of the key secondary endpoint (Section 10.2.1).

Similar analyses will be conducted for the subgroups listed in Section 8.7.

## **10.4. FUNCTIONAL OUTCOMES AND QUALITY OF LIFE ENDPOINTS**

Analyses of functional outcomes and quality of life endpoints will be based on the mITT population.

### **10.4.1. FOSQ-10: Change in the Total Score and 5 Subscales from Baseline to Weeks 1, 4, 8, and 12**

FOSQ-10 total score is the mean of non-missing 5 subscales (General Productivity, Activity Level, Vigilance, Social Outcomes, Intimacy and Sexual Relationship) multiplied by 5..

Change from Baseline in the FOSQ-10 total score to Weeks 1, 4, 8, and 12 will be summarized and analyzed using a similar MMRM model as the primary analysis of co-primary efficacy endpoints (Section 10.1.1). Change from baseline in the 5 FOSQ-10 subscales to Weeks 1, 4, 8, and 12 will also be summarized.

## Statistical Analysis Plan

Similar analyses will be conducted for the subgroups listed in Section 8.7.

### **10.4.2. SF-36v2: Change in 8 Domain Scores, Physical Component Summary (PCS) Score, and Mental Component Summary (MCS) Score from Baseline to Weeks 4, 8, and 12**

The survey is summarized into 8 domains/scales: physical functioning (PF), role-physical (RP), bodily pain (BP), general health (GH), vitality (VT), social functioning (SF), role-emotional (RE), and mental health (MH). From the eight domains, a physical component summary (PCS) (aggregate of the PF; RP; BP and GH scales) and a mental component summary (MCS) (aggregate of the VT; SF; RE and MH scales) measures are derived.

SF-36 items and scales are scored so that a higher score indicates a better health state. Scoring the SF-36 involves several steps. First, 10 items are reverse-coded and an algorithm for an algebraic summation of item scores is applied to produce domain-specific raw scales that account for missing item responses. Domain-specific raw scales are then transformed to a 0 to 100 range, after which a norm-based (T-score) transformation is applied so that each scale ranges from 0 to 100, with a mean of 50 and a standard deviation of 10 in the 1998 general U.S. population (Ware et al 2000). The norm-based transformation is applied so that domain-specific scales can be meaningfully compared between each other.

The mental and physical component summary scores (MCS/PCS) are computed by aggregating domain scores using factor score coefficients from the 1998 general U.S. population (Ware et al 2000). The aggregated summary scores are standardized to have a mean of 50 with a standard deviation of 10 in the general 1998 U.S. population (Ware et al 2000).

When calculating the raw domain scores, if at least half the item scores for a domain are non-missing, the missing item scores will be replaced with the average of the non-missing scores for the domain. Otherwise the raw domain score and corresponding norm-based domain scores will be set to missing.

Domain scores and component summary scores will be calculated using the Quality Metric (QM) Certified Scoring software provided by Optum, Inc.

Changes in domain scores (norm-based), the physical component summary (PCS) score, and the mental component summary (MCS) score from Baseline to Weeks 4, 8 and 12 will be analyzed using a similar MMRM model as the primary analysis of the co-primary efficacy endpoints (Section 10.1.1).

## Statistical Analysis Plan

### 10.4.3. EQ-5D-5L: EQ-5D Dimensions at Week 1, 4, 8, and 12

EQ-5D-5L has 5 dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression). Each dimension has 5 levels of response (e.g. 1=no problem, 2=slight problem, 3=moderate problem, 4=severe problem, 5=unable).

If multiple levels of response are checked for a single dimension, then the level of response will be treated as a “missing value” at the dimension.

The chi-square test will be used to test the treatment differences between each JZP-110 group (JZP-110 dose groups and combined JZP-110) and placebo in the endpoints below.

- Number and percentage of subjects in each of the 5 levels (e.g., no problem, slight problem, moderate problem, severe problem, unable) for each dimension at Baseline and Weeks 1, 4, 8, and 12.
- Number and percentage of subjects reporting any problems (levels 2-5) for each dimension at Baseline and Weeks 1, 4, 8, and 12.

Similar analyses will be conducted for the subgroups listed in Section 8.7.

### 10.4.4. EQ VAS: Change in the VAS Score from Baseline to Weeks 1, 4, 8, and 12

EQ VAS score is ranged from 0 to 100, with a higher score indicating a better health condition. Change in the EQ VAS score from Baseline to Weeks 1, 4, 8, and 12 will be analyzed using a similar MMRM model as the primary analysis of co-primary efficacy endpoints (Section 10.1.1).

EQ VAS score at each time point and change from Baseline to each post-baseline time point will be summarized by treatment group using mean, SD, minimum, 25<sup>th</sup> percentile, median, 75th percentile and maximum.

Similar analyses will be conducted for the subgroups listed in Section 8.7.

### 10.4.5. EQ-5D-5L Index: Index Value from Baseline to Week 1, 4, 8, and 12

Health states in the EQ-5D-5L can be converted into a single index value, where index values are presented in the country specific value sets to facilitate the calculation of quality-adjusted life years (QALYs). EQ-5D-5L value sets can be used to obtain the EQ-5D-5L index values based on the crosswalk for the respective countries: Denmark, France, Germany, Japan, the Netherlands, Spain, Thailand, UK, US, Zimbabwe. References: <http://www.euroqol.org/about-eq-5d/valuation-of-eq-5d/eq-5d-5l-value->

## Statistical Analysis Plan

sets.html; van Hout B, Janssen MF, et al. Interim scoring for the EQ-5D-5L: Mapping the EQ-5D-5L to EQ-5D-3L value sets. Value in Health 2012 Jul-Aug; 15(5):708-15.

Change in the EQ-5D-5L index value from Baseline to Weeks 1, 4, 8, and 12 will be analyzed using a similar MMRM model as the primary analysis of co-primary efficacy endpoints (Section 10.1.1).

EQ-5D-5L index value at each time point and change from Baseline to each post-baseline time point will be summarized by treatment group using mean, SD, minimum, 25th percentile, median, 75th percentile and maximum.

Similar analyses will be conducted for the subgroups listed in Section 8.7.

### **10.4.6. WPAI:SHP: Change in WPAI:SHP Endpoints from Baseline to Week 1,4,8, and 12**

WPAI:SHP includes the following questions:

- 1 = currently employed
- 2a = hours missed from work due to OSA
- 2b = hours missed from work other reasons
- 3 = hours actually worked
- 4 = degree OSA affected productivity while working
- 5 = degree problem affected regular activities

#### **Endpoints and Scoring:**

The following four endpoints will be developed for evaluation of work productivity and activity impairment. Multiply scores by 100 to express in percentages.

- Percent work time missed due to problem:  $\text{Score} = Q2a / (Q2a + Q2b + Q3)$
- Percent impairment while working due to problem:  $\text{Score} = Q4 / 10$
- Percent overall work impairment due to problem:  
 $\text{Score} = (Q2a + Q2b) / (Q2a + Q2b + Q3) + (1 - (Q2a + Q2b) / (Q2a + Q2b + Q3)) \times Q4 / 10$
- Percent activity impairment due to problem:  $\text{Score} = Q5 / 10$

#### **Analyses:**

Change in the WPAI:SHP endpoints below from Baseline to Weeks 1, 4, 8, and 12 will be analyzed using a similar MMRM model as the primary analysis of co-primary efficacy endpoints (Section 10.1.1).

- Percent work time missed due to problem

## Statistical Analysis Plan

- Percent impairment while working due to problem
- Percent overall work impairment due to problem
- Percent activity impairment due to problem

### 10.5. EXPLORATORY ENDPOINTS

#### 10.5.1. Use of Primary OSA Therapy

The use of primary OSA therapy will be summarized by analysis periods, the duration of the analysis period is defined as below:

- Baseline period is the days prior to the day of first dose
- Weeks 1 - 4 period is from the first dose day to the End of Week 4 Visit Day
- Weeks 5 - 8 period is from day after the Week 4 Visit to the End of Week 8 Visit Day
- Weeks 9 - 12 period is from day after the Week 8 Visit to the End of Week 12 Visit Day

If a subject's primary OSA therapy use was missing, the missing daily data will be imputed using LOCF method within the period, up to the early termination date or the date of the end of that period, whichever is earlier. Otherwise, missing data will not be imputed.

The variables below will be summarized for each analysis period by treatment groups:

- Percentage of nights that subjects use primary OSA therapy at Baseline and all post-baseline analysis periods
- Change in the percentage of nights that the primary OSA therapy is used from the Baseline to each post-baseline analysis period

Change in the percentage of nights that the primary OSA therapy is used from Baseline to the each analysis period after Baseline will be analyzed by using a Wilcoxon rank-sum test to compare each treatment group vs placebo and combined JZP-110 vs placebo.

In addition, the following exploratory summary analyses will be conducted:

For subjects who have electronically retrievable data:

- Average number of hours that a subject used OSA device per night in each analysis period, will be calculated as:
  - Total number of hours that a subject used OSA device in a period/ total number of nights that the subject used OSA device in the period. The

## Statistical Analysis Plan

nights that the subject did not use OSA device will not be included in the calculation.

For subjects who do not have electronically retrievable data:

- Percentage of nights that subjects used an OSA device more than half of the night
- Percentage of nights that subjects used an OSA device less than half of the night
- Percentage of nights that subjects used an OSA device with unknown duration

### 10.5.2. Change in PSG Parameters from Baseline to Week 4 and 12

PSG parameters include the following:

- Total sleep time (TST, in minutes)
- Stage N1 time (in minutes)
- Stage N2 time (in minutes)
- Stage N3 time (in minutes)
- Wake time after sleep onset (WASO, in minutes), total minutes of wakefulness recorded after sleep onset, as scored by epochs.
  - $\text{WASO} = \text{Wake epochs} / 2 - (\text{Sleep Latency} \{+ \text{Final Wake Time}\})$
- Number of awakenings
- APNEA Index (AI), the average number of apneas (obstructive, central and mixed) in an hour of sleep
  - Calculated as:  $60 \times \text{total \# of apneas} / \text{total sleep time in minutes}$
- AHI, the average number of apneas and hypopneas in an hour of sleep.
  - Calculated as:  $60 \times \text{total \# of (apneas + hypopneas)} / \text{total sleep time in minutes}$
- Number of central apneas
- Post-baseline minimum SaO2 value
- SaO2

Change in each PSG parameter, except for number of awakenings and number of central apneas, from Baseline to Week 4 and 12 will be analyzed using a similar MMRM model as the primary analysis of co-primary efficacy endpoints (Section 10.1.1).

Change in the number of awakenings and the number of central apneas from Baseline to Week 4 and 12 will be analyzed using a Wilcoxon rank sum test to compare the treatment difference between each JZP-110 group vs placebo and combined JZP-110 vs placebo.

## Statistical Analysis Plan

### 10.5.3 Categorical Analysis for Maintenance of Wakefulness Test

Categorical summaries and accompanying figures will be provided for each treatment group as follows:

- Number and percentage of subjects with an increase in mean MWT sleep latency from Baseline to Week 4 and 12 will be summarized by increase every 5 minutes (e.g.,  $\geq 5$  min,  $\geq 10$  min,  $\geq 15$  min,  $\geq 20$  min,  $\geq 25$  min and  $\geq 30$  min)

### 10.6. SUMMARY OF EFFICACY ANALYSIS METHODS

Endpoint	Time Point	Analysis/ Population	Method	Missing Data Imputation
<b>Co-primary</b>				
Change from Baseline in MWT, ESS	Week 12	Primary/ mITT	MMRM	
		Secondary/ PP	MMRM	
		Sensitivity/ mITT	ANCOVA	SI:LOCF; SI: Mean Imputation;
			MMRM	MI: Multiple Imputation Using MCMC and Regression Method MI: Pattern Mixture Model using Dropout Pattern Imputation
		Subgroup/ mITT	MMRM	
<b>Key Secondary</b>				
Improvement in PGlc	Week 12	Primary/ mITT	Chi-square	SI:LOCF
		Secondary/ PP	Chi-square	
		Sensitivity/ mITT	Chi-square	SI: Varies by early termination reason; SI: Worst-case
		Subgroup/ mITT	Chi-square	SI:LOCF
<b>Other Secondary</b>				



## Statistical Analysis Plan

Endpoint	Time Point	Analysis/ Population	Method	Missing Data Imputation
Change from Baseline in MWT	Week 4	Primary/ mITT	MMRM	
		Subgroup/ mITT	MMRM	
Change from Baseline in ESS	Weeks 1, 4, and 8	Primary/ mITT	MMRM	
		Subgroup/ mITT	MMRM	
Improvement in PGlc	Weeks 1, 4, and 8	Primary/ mITT	Chi-square	SI:LOCF
		Subgroup/ mITT	Chi-square	SI:LOCF
Improvement in CGlc	Weeks 1, 4, 8, and 12	Primary/ mITT	Chi-square	SI:LOCF
		Subgroup/ mITT	Chi-square	SI:LOCF
Time course on MWT	Weeks 4 and 12	Primary / mITT	MMRM	
<b>Functional and Quality of Life</b>				
Change from Baseline in FOSQ-10 total score	Weeks 1, 4, 8, and 12	Primary/ mITT	MMRM	
		Subgroup /mITT	MMRM	
Change from Baseline in SF-36 domain and summary scores	Weeks 4, 8, and 12	Primary/ mITT	MMRM	
EQ-5D-5L Dimensions: number and percentage in each of the 5 levels and reporting any problem (levels 2-5)	Weeks 1, 4, 8, and 12	Primary/ mITT	Chi-square	
		Subgroup/ mITT	Chi-square	
Change from Baseline in EQ VAS	Weeks 1, 4, 8, and 12	Primary/ mITT	MMRM	

## Statistical Analysis Plan

Endpoint	Time Point	Analysis/ Population	Method	Missing Data Imputation
		Subgroup/m ITT	MMRM	
Change from Baseline in EQ-5D-5L Index	Weeks 1, 4, 8, and 12	Primary/ mITT	MMRM	
		Subgroup/m ITT	MMRM	
Change from Baseline in WPAI:SHP scores	Weeks 1, 4, 8, and 12	Primary/ mITT	MMRM	
<b>Exploratory</b>				
Change from Baseline in percentage of use of primary OSA therapy	Analysis Period by every four weeks after Baseline	Primary/ Safety	Wilcoxon rank sum	
Change from Baseline in number of awakenings	Weeks 4 and 12	Primary/ mITT	Wilcoxon rank sum	
Change from Baseline in number of central apneas	Weeks 4 and 12	Primary/ mITT	Wilcoxon rank sum	
Change from Baseline in PSG Parameters, except for number of awakenings and number of central apneas	Weeks 4 and 12	Primary/ mITT	MMRM	

## Statistical Analysis Plan

---

### 11. ANALYSIS OF PHARMACOKINETICS

PK analyses will be performed for the PK Analysis Population.

Concentration of JZP-110, date/time of PK sample collection, date/time of study drug dosing, the flag on whether the study drug was taken on the sampling day and on previous 2 days per schedule and all comments will be listed by nominal sampling time point (Week 1, Week 4, Week 8 Sample 1 and Week 8 Sample 2), sorted by JZP-110 treatment groups. In addition, scatter plot of JZP-110 concentrations over time (Week 1, Week 4, Week 8 Sample 1 and Week 8 Sample 2) will be provided, sorted by JZP-110 treatment groups. Spaghetti plot of each subject's JZP-110 concentration over time (Week 1, Week 4, Week 8 Sample 1 and Week 8 Sample 2) will also be provided, sorted by JZP-110 treatment groups.

Other PK analyses will be included in a separate population PK analysis plan.

## Statistical Analysis Plan

### 12. SAFETY

Safety analyses will be based on the safety population. No inferential statistics will be performed; only summary statistics will be provided unless otherwise noted. Missing safety data will not be imputed.

#### 12.1. EXTENT OF EXPOSURE

Exposure to study drug (in days) is calculated as last dose date of study drug - first dose date of study drug + 1 by treatment group. In addition, overall exposure to JZP-110 (in days) is calculated as last dose date of study drug - first dose date of study drug + 1 regardless of the JZP-110 dose level.

#### 12.2. TREATMENT COMPLIANCE

Study drug compliance (%) is calculated as:  $100 \times (\text{total number of capsules dispensed} - \text{total number of capsules returned}) / \text{total number of capsules expected to be taken during the study (once daily)}$ .

Compliance will be summarized by treatment group. In addition, the number and percentage of subjects in pre-specified compliance categories (< 80%, 80-100%, >100%, >120%) will be summarized by treatment group.

Overall Compliance will also be summarized, regardless of dose level.

#### 12.3. ADVERSE EVENTS

Adverse events will be coded using MedDRA 18 to classify events under primary SOC and PT.

A TEAE is defined as an AE that either begins after the first dose of study drug or worsens after the first dose of study drug.

TEAEs will be summarized by SOC and PT and by treatment groups with sorting based on alphabetical order for the SOC and frequency count (descending order) for the PT. In addition, the incidence of all TEAEs and serious TEAEs will be summarized by PT with sorting based on the frequency count (descending order).

An overview of adverse events will include the number and percent of subjects who had at least one TEAE, Serious TEAE, TEAE related/suspected to be related to study drug, TEAE related/suspected to be related to study procedure, study drug withdrawn due to TEAE, maximum severity of TEAE, or TEAE with fatal outcome.

## Statistical Analysis Plan

Multiple occurrences of an AE are counted only once per subject per SOC and PT for summary tables.

The following TEAEs will be summarized :

- Incidence of all TEAEs
- Incidence of all TEAEs by maximum severity (severe, moderate and mild) specified by investigators
- Incidence of TEAEs related/suspected to be related to study drug specified by investigators
- Incidence of TEAEs related/suspected to be related to study procedure specified by investigators
- Incidence of serious TEAEs
- Incidence of serious TEAEs related/suspected to be related to study drug specified by investigators
- Incidence of TEAEs leading to study drug dose reduction
- Incidence of TEAEs leading to study drug dose increase
- Incidence of TEAEs leading to study drug interruption
- Incidence of TEAEs leading to study drug withdrawn and withdrawn from study
- Incidence of TEAEs in which the outcome is fatal

Incidence of TEAEs occurring in  $\geq 5$  and  $\geq 10\%$  of subjects in any treatment group will be tabulated.

All data collected in the AE case report form (CRF) will be listed in by-subject listings.

### 12.3.1. Subgroup Analyses

The following analyses will be also repeated for the subgroups listed in Section 8.7:

- Overview of AEs
- Incidence of all TEAEs
- Incidence of TEAEs related/suspected to be related to study drug specified by investigators
- Incidence of TEAEs leading to study drug withdrawn and withdrawn from study
- Incidence of Serious TEAEs
- Incidence of TEAEs in which outcome is fatal

### 12.4. LABORATORY EVALUATIONS

If a continuous laboratory value is reported as either below or above the limits of quantification, the qualifiers should be dropped and the numeric value will be used in the analysis (e.g., "< 3" should be "3" and "> 200" should be "200").

## Statistical Analysis Plan

The observed data at each time point and change from baseline at each post-baseline time point in hematology, serum chemistry and quantitative urinalysis test results will be summarized by treatment group.

For hematology and serum chemistry including calculated creatinine clearance, normal ranges for each parameter will be used to categorize the test result as low (value lower than the lower limit), normal (value within the normal range), or high (value higher than the upper limit). For urinalysis, test results will be categorized as normal and abnormal. Frequency counts and percentages will be presented at each time point by treatment groups for these categorical data.

Additional information on clinical significance will also be included in the listing of lab values.

### 12.5. VITAL SIGNS AND 24-HOUR AMBULATORY BLOOD PRESSURE MONITORING

#### 12.5.1. Regular Vital Signs

For the vital signs obtained on the admission to the site in the evenings before the PSG nights at Baseline, Weeks 4 and 12, and for vital signs at Weeks 1, 8, and at the Follow-up visit or Early Termination visit, the observed values at each time point and the change from Baseline in each vital sign parameter to each post-baseline time point will be summarized by treatment group.

The number and percentage of subjects change from Baseline in blood pressure and heart rate to each post baseline time point in the following categories will be summarized:

- % of subjects with an increase in HR of  $\geq 15$  and  $\geq 30$
- % of subjects with a decrease in HR of  $\geq 15$  and  $\geq 30$
- % of subjects with an increase in SBP of  $\geq 10$ ,  $\geq 20$ , and  $\geq 30$
- % of subjects with a decrease in SBP of  $\geq 10$ ,  $\geq 20$ , and  $\geq 30$
- % of subjects with an increase in DBP of  $\geq 10$ ,  $\geq 20$ , and  $\geq 30$
- % of subjects with a decrease in DBP of  $\geq 10$ ,  $\geq 20$ , and  $\geq 30$

Reference ranges for each vital sign parameter will be used to categorize the results as low (lower than the lower limit), within the reference range, or high (higher than the upper limit). In addition, shifts in categories from Baseline to each post-baseline time point for each parameter will be summarized by treatment group.

Reference ranges:

Parameter	Lower Limit of Reference Range	Upper Limit of Reference Range
-----------	--------------------------------	--------------------------------

## Statistical Analysis Plan

Systolic Blood Pressure (mmHg)	90	155
Diastolic Blood Pressure (mmHg)	60	95
Heart Rate (beats/min)	Female: 55, Male 50	Female: 95, Male 90
Respiration Rate (breaths/min)	12	30

### 12.5.2. Blood Pressure and Pulse on the MWT day

On the MWT day at Baseline (Day -1), blood pressure and pulse will be taken approximately 30 minutes after awakening. Blood pressure and pulse will also be taken approximately 2, 3, 5, 7, 9 and 11 hours after awakening.

On MWT days at the end of Weeks 4 and 12 (or Early Termination if PSG/MWT is performed), blood pressure and pulse will be taken approximately 0.5 hour before dosing. Blood pressure and pulse will be taken at approximately 1, 2, 4, 6, 8, and 10 hours after dosing.

The blood pressure and pulse parameter obtained on the MWT day will be summarized as follows:

- Observed values and change from Day -1 (~0.5 hr relative to “Light on”) to Week 4 and 12 (~(-0.5 hr) relative to dosing) visit
- Observed value and changes from the pre-dose (~(-0.5 hr)) to post-dose (~ 1 hr, ~2 hrs, ~4 hrs, ~6 hrs, ~8 hrs, and ~10 hrs) at Week 4 and Week 12 visit and observed value and change from the (~(0.5 hr)) after awakening to the 2, 3, 5, 7, 9 and 11 hours after awakening at baseline (Day -1) visit

In addition, the mean of blood pressure and pulse based on the seven measurements taken during the MWT day (e.g., Day -1 at ~0.5 hr, 2 hrs, 3 hrs, 5 hrs, 7 hr, 9 hrs, and 11 hrs after awakening, Week 4 and 12 at ~(-0.5 hr), ~ 1 hr, ~2 hrs, ~4 hrs, ~6 hrs, ~8 hrs, and ~10 hrs relative to dosing) will be calculated and the mean and mean changes from the Day -1 to the Week 4 and 12 visit will be summarized.

The number and percentage of the subjects with maximum change in blood pressure and pulse from pre-dose to any post-dose time points at the Week 4 and Week 12 visit and the maximum mean change from Day -1 to Week 4 and Week 12 visit will be summarized by following categories:

- % of subjects with an increase in HR of  $\geq 15$  and  $\geq 30$
- % of subjects with a decrease in HR of  $\geq 15$  and  $\geq 30$
- % of subjects with an increase in SBP of  $\geq 10$ ,  $\geq 20$ , and  $\geq 30$
- % of subjects with a decrease in SBP of  $\geq 10$ ,  $\geq 20$ , and  $\geq 30$
- % of subjects with an increase in DBP of  $\geq 10$ ,  $\geq 20$ , and  $\geq 30$

## Statistical Analysis Plan

- % of subjects with a decrease in DBP of  $\geq 10$ ,  $\geq 20$ , and  $\geq 30$

Additional information on clinical significance will also be included in the listing of regular and MWT day vital signs.

### 12.5.3. 24 hour Ambulatory Blood Pressure Monitoring (ABPM)

For the ABPM BP and HR data obtained during Screening and Week 8 at 30 minute interval during a 24 hour period, the overall mean, mean during the daytime period from 7:00 to 22:00 and mean during the nighttime period from 22:00 to 7:00 will be summarized by treatment group.

The following endpoints will also be summarized:

- Change from overall mean value at Screening to overall mean value at Week 8
- Change from daytime mean value (7:00 to 22:00) at Screening to daytime mean value (7:00 to 22:00) at Week 8
- Change from nighttime mean value (22:00 to 7:00) at Screening to nighttime mean value (22:00 to 7:00) at Week 8

In addition, the number and percentage of subjects in the above change from Screening endpoints will also be summarized by the following categories:

- % of subjects with an increase in HR of  $\geq 15$  and  $\geq 30$
- % of subjects with a decrease in HR of  $\geq 15$  and  $\geq 30$
- % of subjects with an increase in SBP of  $\geq 10$ ,  $\geq 20$ , and  $\geq 30$
- % of subjects with a decrease in SBP of  $\geq 10$ ,  $\geq 20$ , and  $\geq 30$
- % of subjects with an increase in DBP of  $\geq 10$ ,  $\geq 20$ , and  $\geq 30$
- % of subjects with a decrease in DBP of  $\geq 10$ ,  $\geq 20$ , and  $\geq 30$

### 12.6. ECG

Observed data at each time point and the change from Baseline to each post-baseline time point in ECG parameters [HR, RR, PR, QRS, QT and QT corrected with Fridericia's formula (QTcF)] will be summarized by treatment group.

The number and percentage of patients with QT and QTcF values falling into the following categories at each post baseline time point will be summarized by treatment groups:

- Change from baseline of 30 - 60 msec in QT and QTcF
- Change from baseline of  $> 60$  msec in QT and QTcF.
- Post-baseline value  $> 480$  msec and baseline value  $\leq 480$  msec in QT and QTcF
- Post-baseline value  $> 500$  msec and baseline value  $\leq 500$  msec in QT and QTcF



## Statistical Analysis Plan

Additional information on clinical significance will also be included in the listing of ECG measurements.

### 12.7. COLUMBIA-SUICIDE SEVERITY RATING SCALE

Suicidal Ideation, Suicidal Behavior, and Self-Injurious Behavior without Suicidal Intent at each time point will be classified by C-SSRS outcomes and composite score.

#### 12.7.1. C-SSRS Outcomes/Composite Scores

The following C-SSRS outcomes have binary responses (yes/no).

- Suicidal Ideation (Categories 1 - 5)
  1. Wish to be dead
  2. Non-specific active suicidal thoughts
  3. Active suicidal ideation with any methods (not plan) without intent to act
  4. Active suicidal ideation with some intent to act, without specific plan
  5. Active suicidal ideation with specific plan and intent
- Suicidal Behavior (Categories 6 - 10)
  6. Preparatory acts or behavior
  7. Aborted attempt
  8. Interrupted attempt
  9. Non-fatal suicide attempt
  10. Completed suicide
- Suicidal Ideation or Behavior (1-10)
- Self-injurious behavior without suicidal intent

Composite endpoints based on the above categories are defined below:

- Suicidal ideation: A “yes” answer during treatment to any one of the five suicidal ideation questions (Categories 1-5) on the C-SSRS.
- Suicidal behavior: A “yes” answer during treatment to any one of the five suicidal behavior questions (Categories 6-10) on the C-SSRS.
- Suicidal ideation or behavior: A “yes” answer during treatment to any one of the ten suicidal ideation and behavior questions (Categories 1-10) on the C-SSRS.

Composite scores based on the above categories are defined below:

Suicidal Ideation score (0 to 5) is based on answers (Yes) to five suicidal ideation questions (Categories 1-5) on the C-SSRS:

## Statistical Analysis Plan

- 
- 0 = No suicidal ideation
  - 1 = Wish to be dead
  - 2 = Non-specific active suicidal thoughts
  - 3 = Activity suicidal ideation with any methods (not plan) without intent to act
  - 4 = Activity suicidal ideation with some intent to act, without specific plan
  - 5 = Activity suicidal ideation with specific plan and intent

Suicidal Behavior score (6 to 10) is based on answers (Yes) to the five suicidal behavior questions (Categories 6-10) on the C-SSRS:

- 0 = No suicidal behavior
- 6 = Preparatory acts or behavior
- 7 = Aborted attempt
- 8 = Interrupted attempt
- 9 = Non-fatal suicide attempt
- 10 = Completed suicide

Suicidal Ideation or Behavior score (0 to 10) is based on answers (Yes) to the ten suicidal ideation and behavior questions (Categories 1-10) on the C-SSRS.

### 12.7.2. Analyses

The number and percentage of subjects having a response of 'Yes' to above outcomes/composite endpoints will be summarized by category and treatment group at each time point.

In addition, to demonstrate treatment emergence in C-SSRS, shift from Baseline (shift to maximum score) to any post-baseline time point in each of the C-SSRS composite scores (Suicidal Ideation, Suicidal Behavior, Suicidal Ideation or Behavior) will be summarized by treatment group. Shift from Baseline to the most serious outcome (No suicidal ideation or behavior, Suicidal Ideation, and Suicidal Behavior) at any post-baseline time point in the C-SSRS outcomes will also be summarized by treatment group.

A listing of subjects with Suicidal Ideation, Suicidal Behavior, or self-injurious behavior without suicidal intent based on the C-SSRS at each time point will be provided.

---

### **13. INTERIM ANALYSES**

No interim analysis is planned.

## Statistical Analysis Plan

### 14. CHANGE FROM ANALYSIS PLANNED IN PROTOCOL

Changes from analysis planned in protocol include the following:

Item	Category	Protocol	SAP
1	Functional outcomes and quality of Life endpoints	SF-36v2: Change in the total score and change in the 8 subscales from Baseline to Week 4, Week 8, and Week 12	SF-36v2: Change in the physical component summary (PCS) score and the mental component summary (MCS) score and change in the 8 subscales from Baseline to Week 4, Week 8, and Week 12
2	Explorative Endpoint	OSA: Change in frequency of use of primary OSA therapy from Baseline to Week 12	Change in the percentage of use of Primary OSA therapy from Baseline to each subsequent analysis period

## Statistical Analysis Plan

---

### 15. REFERENCE

Rubin D.B.: Multiple Imputation for Nonresponse in Surveys. New York: John Wiley and Sons, 1987

Schafer: Analysis of Incomplete Multivariate Data. Chapman & Hall, New York, 1997

Ware JE: SF-36 Health Survey update. Spine, 25(24):3130-3139, 2000

Carpenter, J. R. and Kenward, M. G. (2013), Multiple Imputation and Its Application, New York: John Wiley & Sons.

van Buuren, S. (2012), Flexible Imputation of Missing Data, Boca Raton, FL: Chapman & Hall/CRC.